

# Worst-Case Additive Noise in Wireless Networks

Ilan Shomorony and A. Salman Avestimehr

Cornell University, Ithaca, NY

May 2, 2012

## Abstract

An important classical result in Information Theory states that the Gaussian noise is the worst-case additive noise in point-to-point channels. In this paper, we significantly generalize this result and show that the Gaussian noise is also the worst-case additive noise in general wireless networks with additive noises that are independent from the transmit signals. More specifically, we prove that, given a coding scheme for an AWGN network, one can build a coding scheme that achieves the same rates on an additive noise wireless network with the same topology, where the noise terms may have any distribution with same mean and variance as in the AWGN network.

## 1 Introduction

The modeling of background noise in point-to-point wireless channels as an additive Gaussian noise is well supported from both theoretical and practical viewpoints. In practice, we have witnessed that current wireless systems that were designed based on the assumption of additive Gaussian noise perform quite well. This is intuitively explained by the fact that, from the Central Limit Theorem, the composite effect of many (almost) independent noise sources (e.g., thermal noise, shot noise, etc.) should approach a Gaussian distribution. From a theoretical point of view, Gaussian noise has been proven to be the worst-case noise for additive noise channels. This follows mainly from the fact that the Gaussian distribution maximizes the entropy subject to a variance constraint. More precisely, from the Channel Coding Theorem [1], the capacity of a channel  $f(y|x)$  is given by

$$C = \max_{f(x): E[X^2] \leq P} I(X; Y). \quad (1)$$

Thus, if we choose  $X$  to be distributed as  $\mathcal{N}(0, P)$ , we have that

$$C \geq h(X) - h(X|Y) = \frac{1}{2} \log(2\pi eP) - h(X|Y).$$

In the case of an additive noise (AN) channel  $Y = X + Z$ , where  $E[Z] = 0$  and  $E[Z^2] = \sigma^2$ , the fact that the Gaussian distribution maximizes the entropy implies that  $h(X|Y) \leq \frac{1}{2} \log\left(2\pi e \frac{P\sigma^2}{P+\sigma^2}\right)$ . We conclude that

$$C_{\text{AN}} \geq \frac{1}{2} \log\left(1 + \frac{P}{\sigma^2}\right) = C_{\text{AWGN}},$$

where  $C_{\text{AWGN}}$  is the capacity of the AWGN channel, which is achieved by a Gaussian input distribution. Moreover, a more operational justification of the fact that Gaussian is the worst-case noise for additive

noise channels was provided in [2], where it was shown that random Gaussian codebooks and nearest-neighbor decoding achieve the capacity of the corresponding AWGN channel on a non-Gaussian AN channel.

Once we go beyond point-to-point channels, Gaussian noise is only known to be the worst-case additive noise in some special wireless networks, such as the Multiple Access Channel, the Degraded Broadcast Channel and MIMO channels. In all such cases the capacity has been fully characterized and is known to be achievable with Gaussian inputs. Therefore, similar arguments to the one above can be used to show that, in these cases, Gaussian noise is indeed the worst-case additive noise. However, for more general wireless networks where the capacity is unknown, we lack the tools to make such an assertion. The recent constant-gap capacity approximations for the Interference Channel [3] and for single-source single-destination relay networks [4–6] can only be used to state that Gaussian noise is “approximately” the worst-case additive noise in these cases. Nonetheless, in a leap of faith, most of the research concerning such systems and many other wireless networks views the AWGN channel model as the standard wireless link model. As a result, it remains a fundamental open question whether Gaussian noise is the worst-case additive noise in general wireless networks.

In this work, we answer this question by showing that the Gaussian noise is in fact the worst-case noise for arbitrary wireless networks with additive noises that are independent of the transmit signals. We consider wireless networks with unrestricted topologies and general traffic demands. We show that any coding scheme that achieves a given set of rates on a network with Gaussian additive noises can be used to construct a coding scheme that achieves the same set of rates on a network with same topology and traffic demands, but with non-Gaussian additive noises. It is also important to notice that our coding scheme construction only depends on the mean and variance of the noise distributions of our non-Gaussian network, and is oblivious to their precise statistics. This means that our approach also results in a framework to design codes for networks with unknown noise distributions with an asymptotic performance guarantee.

We prove that the Gaussian noise is the worst-case noise in wireless networks based on two main results. The first one is that, given a coding scheme with *finite reading precision* for an AWGN network, one can build a coding scheme that achieves the same rates on a non-Gaussian wireless network. A coding scheme is said to have finite reading precision if, for any node, its transmit signals only depend on its received signals read up to a finite number of digits after the decimal point. This result is proven in three main steps. We start by applying a transformation at the transmit signals and received signals of all nodes in the network in order to create an “approximately Gaussian” effective network. The technique resembles OFDM in that it uses the Discrete Fourier Transform in order to mix together multiple uses of the same channel. This mixing causes the additive noise terms from distinct network uses to be averaged over time and, by making use of Lindeberg’s Central Limit Theorem [7], it can be shown that the resulting effective noise is approximately Gaussian in the distribution sense. Thus, we create an approximately Gaussian network with dependent noises, since the mixing causes distinct noise realizations at the same receiver to be dependent of each other. The second step is a combination of an interleaving technique and a random outer code, which allows us to handle this dependence among the noise realizations. The interleaving operation creates multiple blocks of network uses inside which the additive noises are i.i.d. and almost normally-distributed. However, in order to be able to apply the original coding scheme that we have for the AWGN network in each of these blocks, we need to make sure that its error probability will not change much when the noise distributions are only approximately Gaussian. This can be done since we require the original coding scheme to have finite reading precision. For such coding schemes, the sets of noise realizations have a special structure and can be shown to be *continuity sets*. It follows from the portmanteau Theorem that the coding scheme’s performance on an almost-Gaussian network does not deviate much from its performance on an actual Gaussian network.

The second main result we need is that, for any wireless network, the capacity when we restrict ourselves to coding schemes with finite reading precision, and allow the precision to tend to infinity along

the sequence of coding schemes, is the same as the unrestricted capacity. To prove this we first show that, for any coding scheme with infinite precision, there exists a quantization scheme of the received signals which does not increase the error probability of the coding scheme too much. This is done by showing that a truncation of the bit expansion of the received signal followed by a random shift performs well; thus, there must exist a fixed shift for each node which guarantees the same performance. This quantization operation makes the coding scheme have finite reading precision, and the result follows.

The paper is organized as follows. In section 2, we describe the network model and introduce the necessary terminology. We start by focusing on wireless networks with  $|L|$  unicast sessions, which makes the proofs simpler and easier to follow. In section 3, we state our main result (Theorem 1) and the two main theorems that are needed for it. Theorem 2 states that coding schemes with finite reading precision can be used to construct coding schemes for non-Gaussian networks. Theorem 3 states that coding schemes with infinite reading precision can be “quantized” yielding coding schemes with finite reading precision that perform almost as well. The proof of Theorem 2 is broken into three different sections as follows. We first describe the OFDM-like scheme in subsection 3.1. Then, in section 3.2, we show that the additive noises obtained from the OFDM-like scheme in fact converge in distribution to Gaussian noises. In section 3.3, we describe the interleaving technique and the outer code that are used to handle the dependence between the noises after the OFDM-like scheme, and we show how the requirement of finite reading precision can be used to show that our coding scheme designed for a Gaussian network can be applied to an almost-Gaussian network without much loss in performance. The proof of Theorem 3 is in section 3.4. In section 4, we describe how we can modify the arguments in the previous sections in order to consider, instead of  $|L|$ -unicast wireless networks, wireless networks with general traffic demands. We conclude the paper in section 5.

## 2 Problem Setup and Definitions

An  $|L|$ -unicast additive noise wireless network  $(G, L)$  consists of a directed graph  $G = (V, E)$ , where  $V$  is the vertex (or node) set and  $E \subseteq V \times V$  is the edge set, and a set  $L \subseteq V \times V$  of source-destination pairs. We assume throughout that all sources and destinations are distinct nodes, although it is possible to strengthen our results to more general settings, as done in section 4. All nodes in  $V$  which are not sources function as relays. We associate a real-valued channel gain  $h_{u,v}$  with each edge  $(u, v) \in E$ .

Communication in a multiple-unicast wireless network is performed over a block of  $n$  discrete time steps. At time  $t = 1, 2, \dots, n$ , each node  $u \in V$  transmits a real-valued signal  $X_u[t]$ , which must satisfy an average power constraint  $\frac{1}{n} \sum_{t=1}^n X_u^2[t] \leq P$ ,  $\forall u \in V$ , for some fixed  $P \geq 0$ . The signal received by node  $v$  at time  $t$  is given by

$$Y_v[t] = \sum_{u \in \mathcal{I}(v)} h_{u,v} X_u[t] + N_v[t], \quad (2)$$

where  $\mathcal{I}(v) = \{u \in V : (u, v) \in E\}$ , and the additive noise  $N_v$  is assumed to be i.i.d. over time and satisfies  $E[N_v] = 0$  and  $E[N_v^2] = \sigma_v^2 < \infty$ . We also assume that the noise terms are independent from all transmit signals and from all noise terms at distinct nodes, and that each  $N_v$  has an absolutely continuous distribution. If all the additive noises in the network are normal  $\mathcal{N}(0, \sigma_v^2)$ , then we say the network is an AWGN network.

**Definition 1.** A coding scheme  $\mathcal{C}$  with block length  $n \in \mathbb{N}$  and rate tuple  $\mathbf{R} = (R_1, \dots, R_{|L|}) \in \mathbb{R}^{|L|}$  for an  $|L|$ -unicast additive noise wireless network consists of:

1. An encoding function  $f_i : \{1, \dots, 2^{nR_i}\} \rightarrow \mathbb{R}^n$  for each source  $s_i$ ,  $i = 1, \dots, |L|$ , where each codeword  $f_i(w_i)$ ,  $w_i \in \{1, \dots, 2^{nR_i}\}$ , satisfies an average power constraint of  $P$ .

2. Relaying functions  $r_v^{(t)} : \mathbb{R}^{t-1} \rightarrow \mathbb{R}$ , for  $t = 1, \dots, n$ , for each relay  $v \in V$  that is not a source, satisfying the average power constraint

$$\frac{1}{n} \sum_{t=1}^n \left[ r_v^{(t)}(y_1, \dots, y_{t-1}) \right]^2 \leq P,$$

for all  $(y_1, \dots, y_{t-1}) \in \mathbb{R}^{t-1}$ .

3. A decoding function  $g_i : \mathbb{R}^n \rightarrow \{1, \dots, 2^{nR_i}\}$  for each destination  $d_i$ ,  $i = 1, \dots, |L|$ .

**Definition 2.** The error probability of a coding scheme  $\mathcal{C}$  (as defined in Definition 1), is given by

$$P_{\text{error}}(\mathcal{C}) = \Pr \left[ \bigcup_{i=1}^{|L|} \{W_i \neq g_i(Y_{d_i}[1], \dots, Y_{d_i}[n])\} \right],$$

where the message transmitted by source  $s_i$ ,  $W_i$ , is assumed to be chosen uniformly at random from  $\{1, \dots, 2^{nR_i}\}$ , for  $i = 1, \dots, |L|$ .

**Definition 3.** A rate tuple  $\mathbf{R}$  is said to be achievable for an  $|L|$ -unicast wireless network  $(G, L)$  if there exists a sequence of coding schemes  $\mathcal{C}_n$  with rate tuple  $\mathbf{R}$  and blocklength  $n$ , for which  $P_{\text{error}}(\mathcal{C}_n) \rightarrow 0$ , as  $n \rightarrow \infty$ . The sequence of coding schemes  $\mathcal{C}_n$ ,  $n = 1, 2, \dots$ , is then said to achieve rate tuple  $\mathbf{R}$ . The capacity region of an  $|L|$ -unicast wireless network is the closure of the set of achievable rate tuples.

We will first focus on coding schemes that have *finite reading precision*. Then we will show that coding schemes with infinite reading precision can be converted into coding schemes with finite reading precision without much loss in performance.

**Definition 4.** A coding scheme  $\mathcal{C}$  is said to have finite reading precision  $\rho \in \mathbb{N}$  if the transmit signal of each (non-source) node  $v$  in the network at each time  $t$  only depends on

$$[Y_v[i]]_\rho \triangleq 2^{-\rho} \lfloor 2^\rho Y_v[i] \rfloor, \text{ for } i = 1, \dots, t-1,$$

as opposed to the complete binary expansion of  $Y_v[i]$ .

**Definition 5.** Rate tuple  $\mathbf{R}$  is achievable by coding schemes with finite reading precision if we have a sequence of coding schemes  $\mathcal{C}_n$ , where coding scheme  $\mathcal{C}_n$  has finite reading precision  $\rho_n$ , which achieves rate tuple  $\mathbf{R}$  according to Definition 3.

*Remark:* Notice that we allow the precision  $\rho_n$  to vary arbitrarily along the sequence of codes, and it may be the case that  $\rho_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

### 3 Main Result

Our main result is to show that any rates that are achievable on a network where each  $N_v$  is normally-distributed for each  $v \in V$  are achieved on a network where each  $N_v$  has an arbitrary absolutely continuous distribution with same mean and variance. More precisely, our main result is the following theorem.

**Theorem 1 (Main Result).** From a sequence of coding schemes that achieve rate tuple  $\mathbf{R}$  on an AWGN  $|L|$ -unicast wireless network  $(G, L)$ , it is possible to construct a single sequence of coding schemes that achieves arbitrarily close to  $\mathbf{R}$  on the same  $|L|$ -unicast wireless network  $(G, L)$ , where, for each relay  $v$ , the distribution of  $N_v$  is replaced with any absolutely continuous distribution satisfying  $E[N_v] = 0$  and

$E[N_v^2] = \sigma_v^2$ . Therefore, if  $C_{\text{AWGN}}$  is the capacity region of the AWGN  $|L|$ -unicast wireless network  $(G, L)$ , and  $C_{\text{non-AWGN}}$  is the capacity region of the same wireless network  $(G, L)$  where, for each relay  $v$ , the distribution of  $N_v$  is replaced with an absolutely continuous distribution satisfying  $E[N_v] = 0$  and  $E[N_v^2] = \sigma_v^2$ , then

$$C_{\text{AWGN}} \subseteq C_{\text{non-AWGN}}.$$

Theorem 1 is proved in the remainder of this section. We will prove it through the following two auxiliary results.

**Theorem 2.** *Suppose a rate tuple  $\mathbf{R}$  is achievable by coding schemes with finite precision on an AWGN wireless network  $(G, L)$ . Then it is possible to construct a single sequence of coding schemes that achieves arbitrarily close to  $\mathbf{R}$  on the same  $|L|$ -unicast additive noise wireless network  $(G, L)$  where, for each relay  $v$ , the distribution of  $N_v$  is replaced with any arbitrary absolutely continuous distribution satisfying  $E[N_v] = 0$  and  $E[N_v^2] = \sigma_v^2$ .*

**Theorem 3.** *Suppose we have a sequence of coding schemes  $C_n$  achieving a rate tuple  $\mathbf{R}$  on a wireless network  $(G, L)$ . Then it is possible to construct a sequence of coding schemes  $C_n^*$  with finite reading precision that also achieves  $\mathbf{R}$  on the same wireless network  $\mathcal{N}$ .*

It is clear that by combining Theorems 2 and 3, Theorem 1 will follow. To prove Theorem 2, we start by assuming that we have a sequence of coding schemes with finite reading precision designed to achieve a rate tuple  $\mathbf{R}$  on an AWGN network. Then, through a series of steps, we will use this sequence of coding schemes to construct another sequence of coding schemes that achieves arbitrarily close to the rate tuple  $\mathbf{R}$  on the corresponding network where the additive noises are not Gaussian.

A diagram illustrating the proof steps of Theorem 2 is shown in Figure 1. We start by describing an OFDM-like scheme that is applied to all nodes in the network. The main idea is that, by applying an Inverse Discrete Fourier Transform (IDFT) to the block of transmit signals of each node, and a Discrete Fourier Transform (DFT) to the block of received signals of each node, we create effective additive noise terms that are weighted averages of the additive noise realizations during that block. We describe this procedure in detail in section 3.1. Then, in section 3.2, we show that this mixture of noises converges in distribution to a Gaussian additive noise term. This is done by showing that the weighted average of the noise realizations satisfies Lindeberg's Central Limit Theorem Condition [7]. Therefore, the OFDM-like scheme effectively produces a network where the noises at each node are dependent across time and approximately Gaussian. The dependence across time is undesirable since our original coding scheme designed for the AWGN network assumed that the additive noise at each receiver is i.i.d. over time. To overcome this problem, in section 3.3, we apply the OFDM-like scheme over multiple blocks, and then we interleave the effective network uses from distinct blocks. This effectively creates several blocks in which the network behaves as an Approximately AWGN network (with i.i.d. noises). Then our original code for the AWGN network can be applied to each approximately AWGN block. The fact that this code has finite reading precision guarantees that, when applied to the approximately AWGN block, its error probability is close to its error probability on the AWGN network. More formally, this means that, for any choice of messages  $\mathbf{w} \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$ , the probability that the joint noise realization  $\mathbf{Z}$  belongs to the error set  $A_{\mathbf{w}}$  (i.e., causes an error to occur) is approximately the same in both cases, which follows from the fact that  $A_{\mathbf{w}}$  can be shown to be a continuity set. Finally, we take care of the dependence between the noises of different blocks created in the interleaving operation by using a random outer code for each source-destination pair. Then we can show via a mutual-information argument that we can achieve a rate tuple arbitrarily close to  $\mathbf{R}$  on the non-Gaussian wireless network.

In section 3.4, we prove Theorem 3. The main idea is to show that, given a coding scheme with infinite reading precision, there exists a set of quantization mappings, one for each node in the network,

such that, if each node quantizes its received signal before applying the relaying or decoding function, the change in the error probability is arbitrarily small.

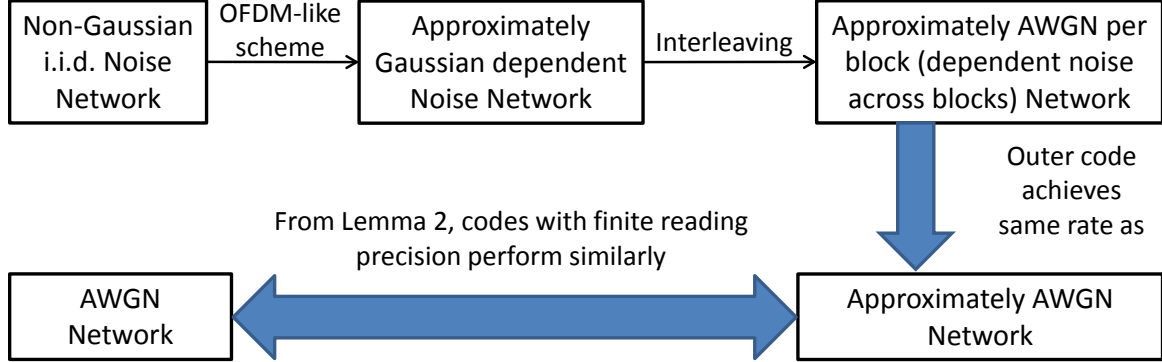


Figure 1: Diagram of proof steps of Theorem 2.

### 3.1 An OFDM-like scheme to mix the noises over time

We use an approach similar to OFDM in order to create an effective network with additive noises that are as close to normally-distributed as we wish. Let us assume that a node  $u \in V$  has  $b$  real-valued signals  $d_0, d_1, \dots, d_{b-1}$  which are the inputs to the effective channels we intend to create. We assume that  $b$  is even, to simplify the expressions. Then node  $u$  “packs” these signals into  $b$  complex numbers  $\tilde{d}_0, \dots, \tilde{d}_{b-1}$  as follows.

$$\begin{aligned}
 \tilde{d}_0 &= d_0 \\
 \tilde{d}_i &= d_{2i-1} + jd_{2i} \quad \text{for } i = 1, \dots, \frac{b}{2} - 1 \\
 \tilde{d}_{b/2} &= d_{b-1} \\
 \tilde{d}_i &= \tilde{d}_{b-i}^* \quad \text{for } i = \frac{b}{2} + 1, \dots, b-1
 \end{aligned}$$

Next, node  $u$  takes IDFT of the vector  $\tilde{\mathbf{d}}_{\mathbf{u}} = (\tilde{d}_0, \dots, \tilde{d}_{b-1})$  to obtain the vector  $\mathbf{X}_{\mathbf{u}} = \text{IDFT}(\tilde{\mathbf{d}}_{\mathbf{u}})$ . Throughout the paper, we assume that DFT and IDFT refer to the *unitary* version of the DFT and IDFT. Since  $\tilde{\mathbf{d}}_{\mathbf{u}}$  is conjugate symmetric,  $\mathbf{X}_{\mathbf{u}}$  is a real vector (in  $\mathbb{R}^b$ ). Moreover, we will require the original real-valued signals to satisfy

$$\text{avg} [d_0^2] \leq P, \tag{3}$$

$$\text{avg} [d_i^2] \leq P/2, \text{ for } i = 1, \dots, b-1, \tag{4}$$

$$\text{avg} [d_b^2] \leq P, \tag{5}$$

where the avg operator refers to time average; i.e., if each  $d_i$  is seen as a stream of signals  $d_i[1], \dots, d_i[k]$ , then  $\text{avg}(d_i) = k^{-1} \sum_{t=1}^k d_i[t]$ . Then we must have, by Parseval’s relationship,

$$\begin{aligned}
 \frac{1}{b} \text{avg} [\|\mathbf{X}_{\mathbf{u}}\|^2] &= \frac{1}{b} \sum_{i=0}^{b-1} \text{avg} [|\tilde{d}_i|^2] \\
 &= \frac{1}{b} \left\{ \text{avg} [d_0^2] + \text{avg} [d_b^2] + 2 \sum_{i=1}^{b/2-1} \text{avg} [d_{2i-1}^2 + d_{2i}^2] \right\} \leq P.
 \end{aligned}$$

Therefore,  $u$  may transmit  $k$  vectors  $\mathbf{X}_u$ , each one over  $b$  time-slots, and the output power constraint over the block  $n = kb$  will be satisfied. A node  $v$  will receive, over each sequence of  $b$  time-slots,

$$\mathbf{Y}_v = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{X}_u + \mathbf{N}_v.$$

By applying a DFT to each block of  $b$  received signals, node  $v$  will obtain

$$\tilde{\mathbf{Y}}_v = \text{DFT}(\mathbf{Y}_v) = \sum_{u \in \mathcal{I}(v)} h_{u,v} \tilde{\mathbf{d}}_u + \text{DFT}(\mathbf{N}_v).$$

This transformation is illustrated in Figure 2.

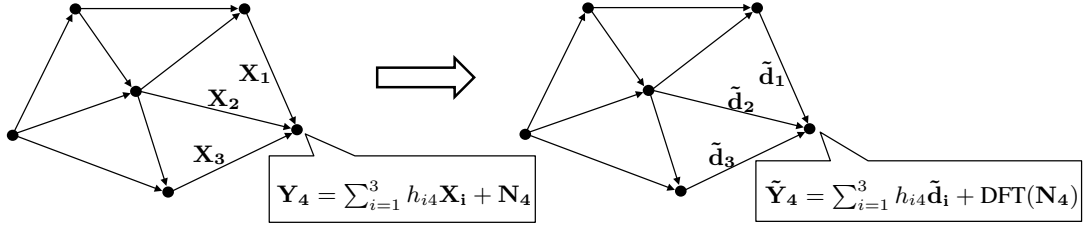


Figure 2: An example of an effective network after OFDM-like scheme.

Next, by looking at each component of  $\tilde{\mathbf{Y}}_v$ , we notice that we have effectively  $b$  complex-valued received signals. The additive noise on the  $\ell^{\text{th}}$  received signal is given by

$$\begin{aligned} \text{DFT}(\mathbf{N}_v)_\ell &= \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} N_v[i] e^{-j2\pi \frac{i\ell}{b}} \\ &= \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} N_v[i] \cos\left(\frac{2\pi i\ell}{b}\right) - j \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} N_v[i] \sin\left(\frac{2\pi i\ell}{b}\right). \end{aligned} \quad (6)$$

By considering the real and imaginary parts of each component  $\tilde{\mathbf{Y}}_{v,i}$  of  $\tilde{\mathbf{Y}}_v$ , for  $i = 0, \dots, b-1$ , separately, we obtain the following  $2b-2$  effective received signals:

- (I)  $\tilde{\mathbf{Y}}_{v,0} = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,0} + \text{DFT}(\mathbf{N}_v)_0$
- (II)  $\Re[\tilde{\mathbf{Y}}_{v,i}] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2i-1} + \Re[\text{DFT}(\mathbf{N}_v)_i] \quad \text{for } i = 1, \dots, \frac{b}{2} - 1$
- (III)  $\Im[\tilde{\mathbf{Y}}_{v,i}] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2i} + \Im[\text{DFT}(\mathbf{N}_v)_i] \quad \text{for } i = 1, \dots, \frac{b}{2} - 1$
- (IV)  $\tilde{\mathbf{Y}}_{v,b/2} = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,b-1} + \text{DFT}(\mathbf{N}_v)_{b/2}$
- (V)  $\Re[\tilde{\mathbf{Y}}_{v,i}] = \sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2(b-i)-1} + \Re[\text{DFT}(\mathbf{N}_v)_i] \quad \text{for } i = \frac{b}{2} + 1, \dots, b-1$
- (VI)  $\Im[\tilde{\mathbf{Y}}_{v,i}] = -\sum_{u \in \mathcal{I}(v)} h_{u,v} \mathbf{d}_{u,2(b-i)} + \Im[\text{DFT}(\mathbf{N}_v)_i] \quad \text{for } i = \frac{b}{2} + 1, \dots, b-1$

However, from the conjugate symmetry of  $\text{DFT}(\mathbf{N}_v)$  (since  $\mathbf{N}_v$  is a real-valued vector), we have that  $\Re[\text{DFT}(\mathbf{N}_v)_i] = \Re[\text{DFT}(\mathbf{N}_v)_{b-i}]$  and  $\Im[\text{DFT}(\mathbf{N}_v)_i] = -\Im[\text{DFT}(\mathbf{N}_v)_{b-i}]$ , for  $i = 1, 2, \dots, b-1$ , and all the received signals from (V) and (VI) are repetitions (up to a change of sign) of the received signals in (II) and (III). Therefore, we conclude that we have effectively  $b$  distinct real-valued received signals with additive noise. It is important to notice that the additive noise terms are *dependent* across these  $b$  received signals.

### 3.2 Noise mixture converges to Gaussian Noise

In this section, we show that the additive noise terms of the effective received signals we obtained in the previous section approximate a Gaussian distribution as  $b$  gets large. We will use the following classical result.

**Theorem 4** (Lindeberg's Central Limit Theorem). *Suppose that for each  $b = 1, 2, \dots$ , the random variables  $Y_{b,1}, Y_{b,2}, \dots, Y_{b,b}$  are independent. In addition, suppose that, for all  $b$  and  $i \leq b$ ,  $E[Y_{b,i}] = 0$ , and let*

$$s_b^2 = \sum_{i=1}^b E[Y_{b,i}^2]. \quad (7)$$

Then, if for all  $\varepsilon > 0$ , Lindeberg's condition

$$\frac{1}{s_b^2} \sum_{i=1}^b E(Y_{b,i}^2 \mathbb{1}_{\{|Y_{b,i}| \geq \varepsilon s_b\}}) \rightarrow 0 \text{ as } b \rightarrow \infty \quad (8)$$

holds, we have that

$$\frac{\sum_{i=1}^b Y_{b,i}}{s_b} \xrightarrow{d} \mathcal{N}(0, 1).$$

Lindeberg's CLT can be used to prove the following lemma.

**Lemma 1.** *Let  $N[0], N[1], N[2], \dots$  be i.i.d. random variables that are zero-mean, have variance  $\sigma^2$  and have an absolutely continuous distribution, and let*

$$Z_b = \frac{1}{\sqrt{b}} \sum_{i=0}^{b-1} N[i] \cos\left(\frac{2\pi i \ell}{b}\right), \quad (9)$$

for some  $\ell \in \{1, \dots, b-1\} \setminus \{b/2\}$ . Then,  $Z_b$  converges in distribution to  $\mathcal{N}(0, \sigma^2/2)$  as  $b \rightarrow \infty$ .

*Proof.* We start by letting  $Y_{b,i+1} = N[i] \cos\left(\frac{2\pi i \ell}{b}\right)$ , for  $i = 0, 1, \dots, b-1$ . Then, by following (7), we have

$$\begin{aligned} s_b^2 &= \sum_{i=1}^b E[Y_{b,i}^2] = \sum_{i=0}^{b-1} E[N[i]^2] \cos^2\left(\frac{2\pi i \ell}{b}\right) \\ &= \frac{\sigma^2}{4} \sum_{i=0}^{b-1} \left(e^{j2\pi \ell \frac{i}{b}} + e^{-j2\pi \ell \frac{i}{b}}\right)^2 = \frac{\sigma^2}{4} \sum_{i=0}^{b-1} \left(e^{j4\pi \ell \frac{i}{b}} + e^{-j4\pi \ell \frac{i}{b}} + 2\right) \\ &= \frac{b\sigma^2}{2} + \frac{\sigma^2}{4} \sum_{i=0}^{b-1} \left(e^{j4\pi \ell \frac{i}{b}} + e^{-j4\pi \ell \frac{i}{b}}\right) = \frac{b\sigma^2}{2} + \frac{\sigma^2(1 - e^{j4\pi \ell})}{4(1 - e^{j4\pi \ell \frac{1}{b}})} + \frac{\sigma^2(1 - e^{-j4\pi \ell})}{4(1 - e^{-j4\pi \ell \frac{1}{b}})} = \frac{b\sigma^2}{2}. \end{aligned}$$

The last equality follows because  $e^{-j4\pi \ell} = 1$  and  $e^{j4\pi \ell \frac{1}{b}} \neq 1$  for any  $\ell \in \{1, \dots, b-1\} \setminus \{b/2\}$ . Next we let  $U_{b,i} = Y_{b,i}^2 \mathbb{1}_{\{|Y_{b,i}| \geq \varepsilon s_b\}} = Y_{b,i}^2 \mathbb{1}_{\left\{|Y_{b,i}| \geq \varepsilon \sigma \sqrt{\frac{b}{2}}\right\}}$ . Consider any sequence  $i_b$ , for  $b = 1, 2, \dots$ , such that  $i_b \in \{1, \dots, b\}$ , and any  $\delta > 0$ . Then we have that

$$\begin{aligned} \Pr(U_{b,i_b} < \delta) &\geq \Pr(|Y_{b,i_b}| < \varepsilon \sigma \sqrt{b/2}) \\ &\geq \Pr(|N[i_b - 1]| < \varepsilon \sigma \sqrt{b/2}) \\ &= \Pr(|N[1]| < \varepsilon \sigma \sqrt{b/2}) \rightarrow 1, \text{ as } b \rightarrow \infty, \end{aligned}$$



which means that  $U_{b,i_b} \xrightarrow{p} 0$  as  $b \rightarrow \infty$ . Moreover, we have that  $|U_{b,i_b}| = U_{b,i_b} \leq N[i_b - 1]^2$  for all  $b$ , and  $E[N[i_b - 1]^2] = \sigma^2 < \infty$ . Therefore, by the Dominated Convergence Theorem, we have that  $E[U_{b,i_b}] \rightarrow 0$  as  $b \rightarrow \infty$ . We conclude that

$$\begin{aligned} \frac{1}{s_b^2} \sum_{i=1}^b E(Y_{b,i}^2 \mathbb{1}_{\{|Y_i| \geq \varepsilon s_b\}}) &= \frac{2}{\sigma^2 b} \sum_{i=1}^b E[U_{b,i}] \\ &\leq \frac{2}{\sigma^2} \max_{1 \leq i \leq b} E[U_{b,i}] \rightarrow 0 \text{ as } b \rightarrow \infty, \end{aligned}$$

and Lindeberg's condition (8) is satisfied for any  $\varepsilon > 0$ . Hence, from Theorem 4, we have that

$$\frac{\sum_{i=1}^b Y_{b,i}}{\sigma \sqrt{b/2}} \xrightarrow{d} \mathcal{N}(0, 1) \implies Z_b = \frac{\sigma}{\sqrt{2}} \frac{\sum_{i=1}^b Y_{b,i}}{\sigma \sqrt{b/2}} \xrightarrow{d} \mathcal{N}(0, \sigma^2/2).$$

□

Now consider the additive noise term in (II). It is the real part of (6), which, by Lemma 1, converges in distribution to  $\mathcal{N}(0, \sigma_v^2/2)$ , as  $b \rightarrow \infty$ . Moreover, it is easy to see that Lemma 1 can be restated with sines replacing the cosines, and the same result will hold. Thus, the additive noise in (III) also converges in distribution to  $\mathcal{N}(0, \sigma_v^2/2)$ . Finally, for the received signals in (I) and (IV), it is easy to see that the additive noise in (6) only has a real component, and by the usual Central Limit Theorem, it converges in distribution to  $\mathcal{N}(0, \sigma_v^2)$ .

### 3.3 Interleaving and Outer Code

In the previous section, we saw that by choosing the length of the OFDM block  $b$  sufficiently large, it is possible to make the effective additive noise at each node  $v$  arbitrarily close (in the distribution sense) to a zero-mean Gaussian noise with variance  $\sigma_v^2/2$  for (II) and (III) and  $\sigma_v^2$  for (I) and (IV). Notice that, since in (4) we restricted the power used in the network uses corresponding to (II) and (III) to  $P/2$ , all of our effective channels have the same SNR they would have if the transmit signals had power  $P$  and the noise variance  $\sigma_v^2$ .

In this section, we address the fact that, as we mentioned before, the additive noise at node  $v$  in the  $b$  effective network uses are dependent of each other. In order to handle this dependence, we consider using the network for a total of  $bk$  times, performing the OFDM-like approach from section 3.1 within each block of  $b$  time steps. Then, by interleaving the symbols, it is possible to view the result as  $b$  blocks of  $k$  network uses. This idea is illustrated in Figure 3. Notice that, within each block of  $k$  network uses, the additive noises are independent, but they are dependent among distinct blocks.

Since from the statement of Theorem 1, the rate tuple  $\mathbf{R}$  is achievable by coding schemes with finite reading precision, we may assume that we have a sequence of coding schemes  $\mathcal{C}_k = (k, \mathbf{R})$  with finite reading precision  $\rho_k$ , whose error probability when used on the AWGN network is

$$\epsilon_k = P_{\text{error}}(\mathcal{C}_k),$$

and satisfies  $\epsilon_k \rightarrow 0$  as  $k \rightarrow \infty$ . Now, consider applying this code over one of the  $b$  blocks of length  $k$  that we obtained from the interleaving. Notice that, in order to apply code  $\mathcal{C}_k$  on a length- $k$  block other than the first or the last one, we will have to divide the output transmit signal of all the nodes by  $\sqrt{2}$  to satisfy (4), but since the additive noises in these blocks have their variance divided by 2 as well, each node can re-scale its received signal by multiplying it by  $\sqrt{2}$ , and the code performs in the exact same way. Now, if  $b$  is chosen fairly large, over this block of length  $k$ , the noises at all nodes are independent and i.i.d. over time, and are very close to Gaussian in distribution, and, intuitively, the error probability we obtain should be close to  $\epsilon_k$ . We will let  $\epsilon_{k,b}$  be the error probability obtained when we apply  $\mathcal{C}_k$

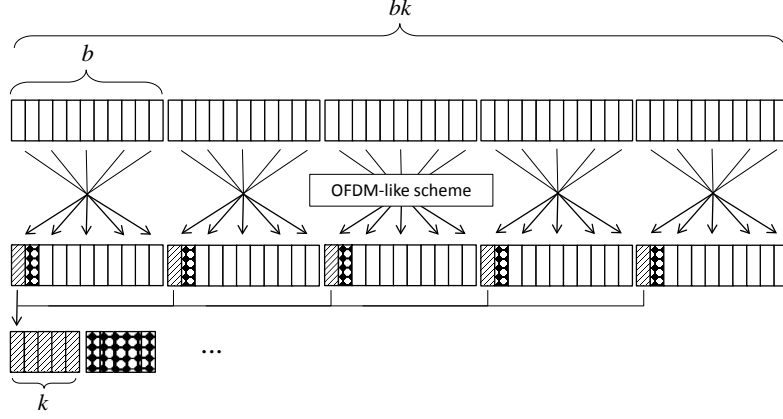


Figure 3: Interleaving the effective network uses obtained from the OFDM-like scheme.

to one of the  $b$  blocks of length  $k$  (notice that this error probability should be the same for any of these blocks).

We let  $\mathbf{Z}_b \in \mathbb{R}^{k|V|}$  be the random vector associated with the effective additive noises at all nodes in  $V$  during this length- $k$  block, assuming that we performed the OFDM-like scheme in blocks of size  $b$ . Since each component of  $\mathbf{Z}_b$  is independent and they all converge in distribution to a zero-mean Gaussian random variable, we have that  $\mathbf{Z}_b$  converges in distribution to a Gaussian random vector. We let  $\mathbf{Z}$  be this limiting distribution, and we know that the component of  $\mathbf{Z}$  corresponding to node  $v$  and time  $\ell$  is distributed as  $\mathcal{N}(0, \sigma_v^2)$  (or  $\mathcal{N}(0, \sigma_v^2/2)$ , depending on the length- $k$  block chosen), for any  $\ell \in \{1, \dots, k\}$ . Now notice that, if we fix the messages chosen at the sources to be  $\mathbf{w} = (w_1, w_2, \dots, w_{|L|}) \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$ , then, whether  $\mathcal{C}_k$  makes an error is only a deterministic function of  $\mathbf{Z}_b$ . Therefore, for each  $\mathbf{w} \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$ , we can define an error set  $A_{\mathbf{w}}$ , corresponding to all realizations of  $\mathbf{Z}_b$  that cause coding scheme  $\mathcal{C}_k$  to make an error. It is important to notice that  $A_{\mathbf{w}}$  is independent of the actual joint distribution of the noise terms; it only depends on the coding scheme  $\mathcal{C}_k$ . Then we can write

$$\epsilon_{k,b} = 2^{-k \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr[\mathbf{Z}_b \in A_{\mathbf{w}}] \quad (10)$$

and also

$$\epsilon_k = 2^{-k \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr[\mathbf{Z} \in A_{\mathbf{w}}]. \quad (11)$$

Our first goal is to show that  $\epsilon_{b,k} \rightarrow \epsilon_k$  as  $b \rightarrow \infty$ . Recall that a Borel set  $A \subseteq \mathbb{R}^m$  is said to be a  $\mu$ -continuity set for some probability measure  $\mu$  on  $\mathbb{R}^m$ , if  $\mu(\partial A) = 0$ , where  $\partial A$  is the boundary of  $A$ . Next, we state the following classical result, which provides an alternative characterization of convergence in distribution.

**Theorem 5** (portmanteau [7]). *Suppose we have a sequence of random vectors  $\mathbf{Z}_b \in \mathbb{R}^{k|V|}$  and another random vector  $\mathbf{Z} \in \mathbb{R}^{k|V|}$ . Let  $\mu_b$  and  $\mu$  be the probability measures on  $\mathbb{R}^{k|V|}$  associated to  $\mathbf{Z}_b$  and  $\mathbf{Z}$  respectively. Then  $\mathbf{Z}_b$  converges in distribution to  $\mathbf{Z}$  if and only if*

$$\lim_{b \rightarrow \infty} \mu_b(A) = \mu(A)$$

for all  $\mu$ -continuity sets  $A$ .

Then, if we let  $\mu$  be the probability measure on  $\mathbb{R}^{k|V|}$  associated to  $\mathbf{Z}$  we have the following Lemma.

**Lemma 2.** Suppose we have a coding scheme  $\mathcal{C} = (k, \mathbf{R})$  with finite reading precision  $\rho$ . Then, for any choice of messages  $\mathbf{w} \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$ , the error set  $A_{\mathbf{w}}$  is a  $\mu$ -continuity set.

*Proof.* Fix some choice of messages  $\mathbf{w}$ . We will use the fact that  $\mathcal{C}$  has finite reading precision  $\rho$  to show that our set  $A_{\mathbf{w}}$  and its complement  $A_{\mathbf{w}}^c = \mathbb{R}^{k|V|} \setminus A_{\mathbf{w}}$  can be represented as a countable union of disjoint convex sets, which will then imply the  $\mu$ -continuity. Recall from Definition 4 that, in a coding scheme with finite reading precision  $\rho$ , a node  $v$  only has access to  $\lfloor Y_v \rfloor_{\rho}$ . Thus, we will call  $\lfloor Y_v \rfloor_{\rho}$  the effective received signal at  $v$ . The set

$$\mathcal{Y} = \left\{ (y_1, \dots, y_{k|V|}) \in \mathbb{R}^{k|V|} : y_i = \lfloor y_i \rfloor_{\rho}, i = 1, \dots, k|V| \right\}$$

can be understood as the set of all possible values of the effective received signals at all nodes in  $V$  during a length- $k$  block. It is clear that  $\mathcal{Y}$  is a countable set for any finite  $\rho$ .

Notice that, for our fixed choice of messages  $\mathbf{w}$ , the vector  $\mathbf{y} \in \mathcal{Y}$  corresponding to the effective received signals at all nodes during the length- $k$  block is a deterministic function of the value of all the noises in the network during the length- $k$  block,  $\mathbf{z} \in \mathbb{R}^{k|V|}$ . Therefore, for each  $\mathbf{y} \in \mathcal{Y}$ , we define  $Q(\mathbf{y}) \subseteq \mathbb{R}^{k|V|}$  to be the set of noise realizations  $\mathbf{z}$  that will result in  $\mathbf{y}$  being the effective received signals. We claim that  $Q(\mathbf{y})$  is a convex set. To see this, consider two noise realizations  $\mathbf{z}, \mathbf{z}' \in Q(\mathbf{y})$  and fix some  $\alpha \in [0, 1]$ . We will show that if we replace one of the components of  $\mathbf{z}$  with the corresponding component of  $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$ , the resulting noise realization  $\mathbf{z}''$  is still in  $Q(\mathbf{y})$ . Then, by using the same argument with  $\mathbf{z}''$  instead of  $\mathbf{z}$ , another component of  $\mathbf{z}''$  is replaced with a component  $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$ , and by repeating this argument, it follows that  $\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}'$  is itself in  $Q(\mathbf{y})$ . So let us focus on the component corresponding to node  $v$  at time  $\ell$ . Let  $y_v[\ell]^*$  be the noiseless version of the received signal at  $v$  at time  $\ell$  with its complete binary expansion. Since  $\mathbf{z}$  and  $\mathbf{z}'$  result in the same  $\mathbf{y}$ , we have that

$$y_v[\ell] = \lfloor y_v[\ell]^* + z_v[\ell] \rfloor_{\rho} = \lfloor y_v[\ell]^* + z'_v[\ell] \rfloor_{\rho}.$$

Now, if we assume wlog that  $z_v[\ell] \leq z'_v[\ell]$ , we have

$$\lfloor y_v[\ell]^* + z_v[\ell] \rfloor_{\rho} \leq \lfloor y_v[\ell]^* + \alpha z_v[\ell] + (1 - \alpha)z'_v[\ell] \rfloor_{\rho} \leq \lfloor y_v[\ell]^* + z'_v[\ell] \rfloor_{\rho}.$$

Thus, it follows that  $y_v[\ell] = \lfloor y_v[\ell]^* + \alpha z_v[\ell] + (1 - \alpha)z'_v[\ell] \rfloor_{\rho}$ , and by replacing  $z_v[\ell]$  with  $\alpha z_v[\ell] + (1 - \alpha)z'_v[\ell]$ , we obtain a noise realization  $\mathbf{z}''$  that is still in  $Q(\mathbf{y})$ , and the claim follows.

In Appendix A, we prove that, for any convex set  $S$ ,  $\lambda(\partial S) = 0$ , where  $\lambda$  is the Lebesgue measure. Moreover, since our measure  $\mu$  is absolutely continuous, it follows by definition that

$$\lambda(S) = 0 \Rightarrow \mu(S) = 0,$$

for any Borel set  $S$ . Thus, since  $\lambda(\partial Q(\mathbf{y})) = 0$ , we have that  $\mu(\partial Q(\mathbf{y})) = 0$ . This, in turn, clearly implies that

$$\mu(Q(\mathbf{y})^\circ) = \mu(\overline{Q(\mathbf{y})}) = \mu(Q(\mathbf{y})), \quad (12)$$

where we use  $S^\circ$  to represent the interior of  $S$  and  $\overline{S}$  to represent its closure. Next, let  $\mathcal{Y}_{A_{\mathbf{w}}} = \{\mathbf{y} \in \mathcal{Y} : A_{\mathbf{w}} \cap Q(\mathbf{y}) \neq \emptyset\}$ . Notice that all noise realizations  $\mathbf{z} \in Q(\mathbf{y})$  will cause all nodes and, in particular, the destination nodes to effectively receive the exact same signals. Therefore, it must be the case that, if  $A_{\mathbf{w}} \cap Q(\mathbf{y}) \neq \emptyset$ , then  $Q(\mathbf{y}) \subseteq A_{\mathbf{w}}$ , which implies that

$$\bigcup_{\mathbf{y} \in \mathcal{Y}_{A_{\mathbf{w}}}} Q(\mathbf{y}) = A_{\mathbf{w}}.$$

Moreover, it is obvious that any noise realization must belong to exactly one set  $Q(\mathbf{y})$ , and we have

$$\bigcup_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}_{A_{\mathbf{w}}}} Q(\mathbf{y}) = A_{\mathbf{w}}^c.$$

Finally, we obtain

$$\begin{aligned} \mu(A_{\mathbf{w}}^\circ) &\stackrel{(i)}{\geq} \mu\left(\bigcup_{\mathbf{y} \in \mathcal{Y}_{A_{\mathbf{w}}}} Q(\mathbf{y})^\circ\right) \\ &\stackrel{(ii)}{=} \sum_{\mathbf{y} \in \mathcal{Y}_{A_{\mathbf{w}}}} \mu(Q(\mathbf{y})^\circ) \stackrel{(iii)}{=} \sum_{\mathbf{y} \in \mathcal{Y}_{A_{\mathbf{w}}}} \mu(Q(\mathbf{y})) \\ &= 1 - \sum_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}_{A_{\mathbf{w}}}} \mu(Q(\mathbf{y})) = 1 - \sum_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}_{A_{\mathbf{w}}}} \mu(Q(\mathbf{y})^\circ) \\ &= 1 - \mu\left(\bigcup_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}_{A_{\mathbf{w}}}} Q(\mathbf{y})^\circ\right) \geq 1 - \mu((A_{\mathbf{w}}^c)^\circ) \\ &= \mu(((A_{\mathbf{w}}^c)^\circ)^c) = \mu(\overline{A_{\mathbf{w}}}), \end{aligned}$$

where (i) follows since, for sets  $B_1, B_2, \dots, (\cup_i B_i)^\circ \supseteq \cup_i B_i^\circ$ , (ii) follows from the countability of  $\mathcal{Y}_{A_{\mathbf{w}}}$  and the fact that  $Q(\mathbf{y}_1) \cap Q(\mathbf{y}_2) = \emptyset$  for  $\mathbf{y}_1 \neq \mathbf{y}_2$ , and (iii) follows from (12). We conclude that  $\mu(\partial A_{\mathbf{w}}) = \mu(\overline{A_{\mathbf{w}}}) - \mu(A_{\mathbf{w}}^\circ) = 0$ .  $\square$

Now it follows from Theorem 5 and Lemma 2 that, for all message choices  $\mathbf{w}$ , we will have

$$\lim_{b \rightarrow \infty} \Pr[\mathbf{Z}_b \in A_{\mathbf{w}}] = \Pr[\mathbf{Z} \in A_{\mathbf{w}}], \quad (13)$$

which implies that  $\epsilon_{b,k} \rightarrow \epsilon_k$  as  $b \rightarrow \infty$ .

From the previous discussion, we see that we can apply code  $\mathcal{C}_k$  within each of the  $b$  blocks of length  $k$  and obtain a probability of error (within that block) that tends to  $\epsilon_k$  as  $b \rightarrow \infty$ . However, since we have a total of  $b$  blocks of length  $k$ , we make an error if we make an error in any of the  $b$  blocks of length  $k$ . It turns out that a simple union bound does not work here, since the error probability would be of the form  $b\epsilon_{b,k}$  and we would not be able to guarantee that it tends to 0 as  $b$  and  $k$  go to infinity. Instead we consider using an outer code for each source-destination pair.

The idea is to apply coding scheme  $\mathcal{C}_k$  to each of the  $b$  length- $k$  blocks, and then view this as creating a discrete channel for each source-destination pair. More specifically, for each length- $b$  block, source  $s_j$  chooses a *symbol* (rather than a message) from  $\{1, \dots, 2^{kR_j}\}^b$  and transmits the  $b$  corresponding codewords from  $\mathcal{C}_k$ . Then destination  $d_j$  will apply the decoder from code  $\mathcal{C}_k$  inside each length- $k$  block and obtain an output symbol also from  $\{1, \dots, 2^{kR_j}\}^b$ . Notice that, by viewing the input to  $b$  network uses as a single input to this discrete channel, we make sure we have a discrete *memoryless* channel, and we can use the Channel Coding Theorem. We can view  $W_j^b$  and  $\hat{W}_j^b$  as the discrete input and output of the channel between  $s_j$  and  $d_j$ . We will then construct a code (whose rate is to be determined) for this discrete channel between  $s_j$  and  $d_j$  by picking each entry uniformly at random from  $\{1, \dots, 2^{kR_j}\}^b$ .

Then, source-destination pair  $(s_j, d_j)$  can achieve rate

$$\begin{aligned}
\frac{1}{bk} I(W_j^b; \hat{W}_j^b) &= \frac{1}{bk} \left( H(W_j^b) - H(W_j^b | \hat{W}_j^b) \right) \\
&\geq R_j - \frac{1}{bk} \sum_{i=1}^b H(W_j[i] | \hat{W}_j[i]) \\
&\stackrel{(i)}{\geq} R_j - \frac{1}{k} (1 + \epsilon_{b,k} k R_j) \\
&= R_j (1 - \epsilon_{b,k}) - \frac{1}{k},
\end{aligned}$$

where (i) follows from Fano's Inequality, since, within each length- $k$  block, we are applying code  $\mathcal{C}_k$  and we have an average error probability of at most  $\epsilon_{b,k}$  (it should in fact be much less than  $\epsilon_{b,k}$  since we are only considering the error event  $W_j[i] \neq \hat{W}_j[i]$  and  $\epsilon_{b,k}$  refers to the union of these events for all source-destination pairs).

We conclude that, by choosing  $b$  and  $k$  sufficiently large, it is possible for each source-destination pair to achieve arbitrarily close to rate  $R_j$ . Thus, our coding scheme can achieve arbitrarily close to the rate tuple  $\mathbf{R}$ . This concludes the proof of Theorem 2.

### 3.4 Optimality of Coding Schemes with Finite Reading Precision

In this section, we prove Theorem 3. This theorem implies that, if we restrict ourselves to coding schemes with finite reading precision, and allow the reading precision to tend to infinity along the sequence of coding schemes, we can achieve any point in the capacity region of a wireless network, thus characterizing the optimality of coding schemes with finite reading precision. We start by considering a sequence of coding schemes  $\mathcal{C}_n$  (with infinite reading precision) that achieves rate tuple  $\mathbf{R}$  on a  $|L|$ -unicast wireless network  $(G, L)$ . Then we will build a sequence of coding schemes  $\mathcal{C}_n^*$  with finite reading precision that also achieves rate tuple  $\mathbf{R}$  on  $(G, L)$ .

Let  $\epsilon_n$  be the error probability of coding scheme  $\mathcal{C}_n$ , which achieves rate tuple  $\mathbf{R}$  on  $(G, L)$ . From Definition 3, we have that  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . For any fixed  $n$ , we will first build a sequence of coding schemes with finite reading precision  $\mathcal{C}_{m,n}^*$ ,  $m = 1, 2, \dots$ , for  $(G, L)$ , such that code  $\mathcal{C}_{m,n}^*$  has error probability  $\epsilon_{m,n}$ , where  $\epsilon_{m,n} \rightarrow \epsilon_n$  as  $m \rightarrow \infty$ . This will then allow us to choose a finite  $m$  for which  $\epsilon_{m,n}$  is arbitrarily close to  $\epsilon_n$ . Notice that, from Definition 1, relaying and decoding functions should be deterministic. However, in order to construct coding scheme  $\mathcal{C}_{m,n}^*$ , we will first assume that the relaying and decoding functions are allowed to be randomized, and later we will derandomize the constructed coding scheme. Recall that, from Definition 1, coding scheme  $\mathcal{C}_n$  is comprised of encoding functions  $\{f_i : 1 \leq i \leq |L|\}$ , relaying functions  $\{r_v^{(t)} : v \in V, 1 \leq t \leq n\}$  and decoding functions  $\{g_i : 1 \leq i \leq |L|\}$ . We will build  $\mathcal{C}_{m,n}^*$  from  $\mathcal{C}_n$  by using the same encoding functions  $f_i$ ,  $i = 1, \dots, |L|$ , and replacing the relaying functions with

$$\tilde{r}_v^{(t)}(Y_v[1], \dots, Y_v[t-1]) \triangleq r_v^{(t)}(\tilde{Y}_v^{(m)}[1], \dots, \tilde{Y}_v^{(m)}[t-1])$$

for  $1 \leq t \leq n$  and  $v \in V$ , and replacing the decoding functions with

$$\tilde{g}_i(Y_v[1], \dots, Y_v[n]) \triangleq g_i(\tilde{Y}_v^{(m)}[1], \dots, \tilde{Y}_v^{(m)}[n]),$$

for  $1 \leq i \leq |L|$ , where we define

$$\tilde{Y}_v^{(m)}[t] = \lfloor Y_v[t] \rfloor_m + U_v^{(m)}[t], \tag{14}$$

for  $v \in V$  and  $1 \leq t \leq n$ , where  $U_v^{(m)}[1], \dots, U_v^{(m)}[n]$  are independent uniform random variables drawn from  $(-2^{-m-1}, 2^{-m-1})$ , independent from all signals and noises in the network. Notice that, since the relaying functions  $r_v^{(t)}$  satisfy the power constraint in Definition 1, so will the new relaying functions  $\tilde{r}_v^{(t)}$ . In order to relate the error probability of  $\mathcal{C}_{m,n}^*$  to the error probability of  $\mathcal{C}_n$ , we will need the following lemma, whose proof is in the Appendix.

**Lemma 3.** *Suppose  $Y$  is a random variable with density  $f$ . Let  $\tilde{Y}^{(m)} = \lfloor Y \rfloor_m + U^{(m)}$ , where  $U^{(m)}$  is uniformly distributed in  $(-2^{-m-1}, 2^{-m-1})$  and independent from  $Y$ . Then each  $\tilde{Y}^{(m)}$  has a density  $f^{(m)}$ , and  $f^{(m)}$  converges pointwise almost everywhere to  $f$ .*

This lemma will be used to show that, by picking  $m$  sufficiently large, we can make the error probability of code  $\mathcal{C}_{m,n}^*$  arbitrarily close to  $\epsilon_n$ . Suppose we fix the message vector  $\mathbf{w} \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$  and let  $\mathbf{Y}$  be the random vector of length  $n|V|$  corresponding to all the received signals at all nodes during the  $n$  time steps in the block if code  $\mathcal{C}_n$  is used. More precisely, we write  $\mathbf{Y} = (\mathbf{Y}[1], \dots, \mathbf{Y}[n])$ , where  $\mathbf{Y}[t] = (Y_1[t], \dots, Y_{|V|}[t])$  is the random vector of received signals at all  $|V|$  nodes at time  $t$ , for  $1 \leq t \leq n$ . The received signal at node  $v$  at time  $t$ ,  $Y_v[t]$ , is defined in (2). Notice that here we assume that the set of nodes  $V$  can be written as  $V = \{1, \dots, |V|\}$ , in order to simplify some expressions. We claim that the random vector  $\mathbf{Y}$  conditioned on the choice of messages  $\mathbf{W} = \mathbf{w}$  has a density. To see this, we notice that, conditioned on the received signals received up to time  $t-1$ , i.e., on  $(\mathbf{Y}[1], \dots, \mathbf{Y}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$ , and on  $\mathbf{W} = \mathbf{w}$ , the transmit signals at time  $t$ ,  $X_v[t]$  for  $v \in V$ , are all deterministic. Thus, the received signals  $Y_v[t]$ , for  $v \in V$ , are conditionally independent and each one is normally-distributed, conditioned on  $(\mathbf{Y}[1], \dots, \mathbf{Y}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$  and  $\mathbf{W} = \mathbf{w}$ . Therefore, the conditional pdf  $f_{Y_v[t]|\mathbf{Y}[1], \dots, \mathbf{Y}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w})$  exists for each  $v \in V$ . We conclude that, conditioned on  $\mathbf{W} = \mathbf{w}$ , the random vector  $\mathbf{Y}$  has a density given by

$$f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{v=1}^{|V|} f_{Y_v[1]|\mathbf{W}}(y_v[1]|\mathbf{w}) \prod_{t=2}^n \prod_{v=1}^{|V|} f_{Y_v[t]|\mathbf{Y}[1], \dots, \mathbf{Y}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w}). \quad (15)$$

Similarly, we let  $\tilde{\mathbf{Y}}^{(m)}$  be the vector of  $n|V|$  effective received signals (14) if code  $\mathcal{C}_{m,n}^*$  is used instead, i.e.,  $\tilde{\mathbf{Y}}^{(m)} = (\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[n])$ , where  $\tilde{\mathbf{Y}}[t] = (\tilde{Y}_1^{(m)}[t], \dots, \tilde{Y}_{|V|}^{(m)}[t])$ . Then, when we condition on  $(\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$ , and on  $\mathbf{W} = \mathbf{w}$ , the transmit signals at time  $t$ ,  $X_v[t]$  for  $v \in V$ , are all deterministic, and the effective received signals  $\tilde{Y}_v^{(m)}[t]$ , for  $v \in V$ , are conditionally independent (although not normally-distributed). To see that the conditional pdf  $f_{\tilde{Y}_v^{(m)}[t]|\mathbf{Y}[1], \dots, \mathbf{Y}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w})$  exists, we notice that, from (14),  $\tilde{Y}_v^{(m)}[t]$  is the sum of two independent random variables (even when conditioned on  $(\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$  and  $\mathbf{W} = \mathbf{w}$ ), and, since  $U_v^{(m)}[t]$  has a density, so does  $\tilde{\mathbf{Y}}^{(m)}[t]$  (see page 266 in [8]). Therefore, we can write the conditional pdf of  $\tilde{\mathbf{Y}}^{(m)}$  conditioned on  $\mathbf{W}$  as

$$f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{v=1}^{|V|} f_{\tilde{Y}_v^{(m)}[1]|\mathbf{W}}(y_v[1]|\mathbf{w}) \prod_{t=2}^n \prod_{v=1}^{|V|} f_{\tilde{Y}_v^{(m)}[t]|\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w}). \quad (16)$$

As we mentioned before, conditioned on  $(\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$  and  $\mathbf{W} = \mathbf{w}$  (or just  $\mathbf{W} = \mathbf{w}$ , if  $t = 1$ ),  $Y_v[t]$  is normally-distributed and thus has a density. Therefore, the random

variables  $\tilde{Y}_v^{(m)} = \lfloor Y_v[t] \rfloor_m + U_v^{(m)}[t]$ , for  $m = 1, 2, \dots$ , conditioned on  $(\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1]) = (\mathbf{y}[1], \dots, \mathbf{y}[t-1])$  and  $\mathbf{W} = \mathbf{w}$ , satisfy the conditions of Lemma 3, and we have that

$$\begin{aligned} f_{\tilde{Y}_v^{(m)}[1]|\mathbf{W}}(y_v[1]|\mathbf{w}) &\rightarrow f_{Y_v[1]|\mathbf{W}}(y_v[1]|\mathbf{w}) \quad \text{and} \\ f_{\tilde{Y}_v^{(m)}[t]|\tilde{\mathbf{Y}}^{(m)}[1], \dots, \tilde{\mathbf{Y}}^{(m)}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w}) &\rightarrow \\ f_{Y_v[t]|\mathbf{Y}[1], \dots, \mathbf{Y}[t-1], \mathbf{W}}(y_v[t]|\mathbf{y}[1], \dots, \mathbf{y}[t-1], \mathbf{w}), \end{aligned}$$

as  $m \rightarrow \infty$ , for  $t = 2, \dots, n$  and  $v \in V$ , for almost all  $\mathbf{y} \in \mathbb{R}^{n|V|}$ . Therefore, we conclude that  $f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) \rightarrow f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w})$  as  $m \rightarrow \infty$  for almost all  $\mathbf{y} \in \mathbb{R}^{n|V|}$  and any  $\mathbf{w} \in \prod_{i=1}^{|L|} \{1, \dots, 2^{kR_i}\}$ .

Next we notice that, conditioned on the message vector  $\mathbf{W} = \mathbf{w}$ , whether we make an error or not is a function of the received signals at all nodes during the  $n$  time steps (it is actually only a function of the received signals at the destinations, but that is irrelevant). Thus, there exists a set  $E_{\mathbf{w}} \subseteq \mathbb{R}^{n|V|}$  of received signals during the  $n$  time steps which cause a decoding error (at any of the decoders). We will let  $\mu_{\mathbf{w}}^{(n)}$  be the probability measure on  $\mathbb{R}^{n|V|}$  corresponding to  $\mathbf{Y}$  (the received signals when using coding scheme  $\mathcal{C}_n$ ) conditioned on  $\mathbf{W} = \mathbf{w}$  and  $\mu_{\mathbf{w}}^{(m,n)}$  be the probability measure on  $\mathbb{R}^{n|V|}$  corresponding to  $\tilde{\mathbf{Y}}^{(m)}$  (the effective received signals when we use coding scheme  $\mathcal{C}_{m,n}^*$ ) conditioned on  $\mathbf{W} = \mathbf{w}$ . By Scheffé's Theorem [8], we have that

$$\sup_{A \in \mathcal{B}} \left| \mu_{\mathbf{w}}^{(n)}(A) - \mu_{\mathbf{w}}^{(m,n)}(A) \right| \leq \int_{\mathbb{R}^{n|V|}} \left| f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) - f_{\tilde{\mathbf{Y}}^{(m)}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) \right| d\lambda \rightarrow 0, \text{ as } m \rightarrow \infty,$$

where  $\mathcal{B}$  is the Borel  $\sigma$ -field on  $\mathbb{R}^{n|V|}$ , and  $\lambda$  is the Lebesgue measure. This, in turn, implies that for any choice of messages  $\mathbf{w}$ , we must have  $\lim_{m \rightarrow \infty} \mu_{\mathbf{w}}^{(m,n)}(E_{\mathbf{w}}) = \mu_{\mathbf{w}}^{(n)}(E_{\mathbf{w}})$ . We conclude that

$$\epsilon_{m,n} = 2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr \left[ \tilde{\mathbf{Y}}^{(m)} \in E_{\mathbf{w}} \mid \mathbf{W} = \mathbf{w} \right] \quad (17)$$

$$= 2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \mu_{\mathbf{w}}^{(m,n)}(E_{\mathbf{w}}) \xrightarrow{m \rightarrow \infty} 2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \mu_{\mathbf{w}}^{(n)}(E_{\mathbf{w}}) = \epsilon_n. \quad (18)$$

Therefore, we can choose, for each  $n$ ,  $m_n$  sufficiently large such that the probability of error of code  $\mathcal{C}_{m_n,n}^*$ ,  $\epsilon_{m_n,n}$ , is at most  $2\epsilon_n$ . Finally, we need to take care of the fact that  $\mathcal{C}_{m_n,n}^*$  uses randomized relaying and decoding functions. First, we notice that if we let  $\mathbf{U}_m$  be the random vector corresponding to the  $n|V|$  samples from  $U(-2^{-(m+1)}, 2^{-(m+1)})$  used by the  $|V|$  nodes during  $n$  time steps, then we can write

$$\begin{aligned} \epsilon_{m_n,n} &= 2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr \left[ \tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \mid \mathbf{W} = \mathbf{w} \right] \\ &= E \left[ 2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr \left[ \tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \mid \mathbf{W} = \mathbf{w}, \mathbf{U}_{m_n} \right] \right]. \end{aligned}$$

Therefore, there must exist some  $\mathbf{u} \in \mathbb{R}^{n|V|}$  for which

$$2^{-n \sum_{i=1}^{|L|} R_i} \sum_{\mathbf{w}} \Pr \left[ \tilde{\mathbf{Y}}^{(m_n)} \in E_{\mathbf{w}} \mid \mathbf{W} = \mathbf{w}, \mathbf{U}_{m_n} = \mathbf{u} \right] \leq \epsilon_{m_n,n}.$$

Thus, we define  $\mathcal{C}_n^*$  to be the coding scheme by having each node  $v$  at time  $t$  quantize its received signal with resolution  $m_n$ , add to it  $u_v[t]$  (i.e., the entry of  $\mathbf{u}$  corresponding to node  $v$  and time  $t$ ) and then apply the relaying/decoding function from code  $\mathcal{C}_n$ . It is then clear that  $\mathcal{C}_n^*$  has deterministic relaying/decoding functions, and its error probability is at most  $\epsilon_{m_n,n} \leq 2\epsilon_n$ . Therefore, the sequence of codes  $\mathcal{C}_n^*$ ,  $n = 1, 2, \dots$ , has finite reading precision and achieves the rate tuple  $\mathbf{R}$ .

## 4 Extension to General Traffic Demands

One immediate extension of the result in Theorem 1 is to consider wireless networks with general traffic demands. These could include non-unicast flows such as multicast and broadcast flows. We again consider an additive noise wireless network described by a directed graph  $G = (V, E)$ . This time, however, each node  $v \in V$  is a source and has a message  $w(v, D)$  for each set of destinations  $D \in \mathcal{P}(V)$ , where  $\mathcal{P}(V)$  is the power set of  $V$ . We can now replace Definition 1 with the following.

**Definition 6.** A coding scheme  $\mathcal{C}$  with block length  $n \in \mathbb{N}$  and rate tuple  $\mathbf{R} \in \mathbb{R}^{V \times \mathcal{P}(V)}$  for an additive noise wireless network consists of:

1. Encoding/relaying functions  $r_v^{(t)} : \mathbb{R}^{t-1} \times \prod_{D \in \mathcal{P}(V)} \{1, \dots, 2^{nR(v,D)}\} \rightarrow \mathbb{R}$ , for  $t = 1, \dots, n$ , for each node  $v \in V$ , satisfying the average power constraint

$$\frac{1}{n} \sum_{t=1}^n \left[ r_v^{(t)}(y_1, \dots, y_{t-1}, \mathbf{w}_v) \right]^2 \leq P,$$

for all  $(y_1, \dots, y_{t-1}) \in \mathbb{R}^{t-1}$  and  $\mathbf{w}_v \in \prod_{D \in \mathcal{P}(V)} \{1, \dots, 2^{nR(v,D)}\}$ .

2. A decoding function  $g_u : \mathbb{R}^n \rightarrow \prod_{\substack{v \in V, \\ D \in \mathcal{P}(V): u \in D}} \{1, \dots, 2^{nR(v,D)}\}$  for each node  $u \in V$ .

With this definition of a coding scheme, it is straightforward to extend Definitions 3, 4 and 5. Then we generalize Theorem 1 as follows.

**Theorem 6.** Suppose a rate tuple  $\mathbf{R}$  is achievable on an AWGN wireless network  $G$ . Then it is possible to construct a sequence of coding schemes that achieves arbitrarily close to  $\mathbf{R}$  on the same additive noise wireless network  $G$  where, for each relay  $v$ , the distribution of  $N_v$  is replaced with an arbitrary absolutely continuous distribution satisfying  $E[N_v] = 0$  and  $E[N_v^2] = \sigma_v^2$ .

Theorem 6 can be proved using the same steps in the proof of Theorem 1. To re-prove Theorem 2 in this new setting, we start by applying the OFDM-like scheme to the transmit and received signals of every node exactly as done in section 3.1. Thus, the convergence in distribution of the effective additive noise terms to Gaussian, proved in section 3.2, still holds. Therefore, we may assume that, as in the beginning of section 3.3, we have  $k$  blocks of  $b$  network uses each, and we apply the OFDM-like scheme inside each length- $b$  block. Next, by interleaving the network uses, we obtain  $b$  blocks of length  $k$  inside which the network is approximately AWGN. Furthermore, since we start off with a sequence of coding schemes  $\mathcal{C}_k = (k, \mathbf{R})$  with finite reading precision, the proof of Lemma 2 holds verbatim, except that  $\mathbf{w}$ , the vector of messages chosen, is now a vector in  $\prod_{v \in V, D \in \mathcal{P}(V)} \{1, \dots, 2^{nR(v,D)}\}$ . Thus, within each length- $k$  block, the probability that any node decodes any of its messages incorrectly (assuming all messages are chosen independently and uniformly at random) is given by  $\epsilon_{b,k}$ , where  $\epsilon_{b,k} \rightarrow \epsilon_k$  as  $b \rightarrow \infty$  and  $\epsilon_k \rightarrow 0$  as  $k \rightarrow \infty$ .

In order to deal with the dependence between the noise realizations of different length- $k$  blocks, we will again consider employing outer codes. This time, however, instead of having one outer code for each source-destination pair, we will have one outer code for each message  $w(v, D)$  (i.e., one outer code for each  $v \in V$  and  $D \in \mathcal{P}(V)$ ). Thus, for each  $v \in V$  and  $D \in \mathcal{P}(V)$ , we will define a *broadcast* discrete channel with input and output alphabet  $\{1, \dots, 2^{kR(v,D)}\}^b$ , where  $v$  is the source and all nodes in  $D$  are the destinations, which are all interested in the same message. We construct each code by sampling  $\{1, \dots, 2^{kR(v,D)}\}^b$  uniformly at random. Let  $W(v, d)^b$  correspond to a random symbol chosen by  $v$  uniformly at random from  $\{1, \dots, 2^{kR(v,D)}\}^b$ , and  $\hat{W}(v, d)_u^b$  be the corresponding output symbol at



each node  $u \in D$ . For the outer code associated with  $v$  and  $D$ , we can achieve rate

$$\begin{aligned}
\frac{1}{bk} \min_{u \in D} I(W(v, D)^b; \hat{W}(v, D)_u^b) &= \frac{1}{bk} \min_{u \in D} \left( H(W(v, D)^b) - H(W(v, D)^b | \hat{W}(v, D)_u^b) \right) \\
&\geq R(v, D) - \max_{u \in D} \frac{1}{bk} \sum_{i=1}^b H(W(v, D)[i] | \hat{W}(v, D)_u[i]) \\
&\stackrel{(i)}{\geq} R(v, D) - \max_{u \in D} \frac{1}{k} (1 + \epsilon_{b,k} k R(v, D)) \\
&= R(v, D)(1 - \epsilon_{b,k}) - \frac{1}{k},
\end{aligned}$$

where (i) follows from Fano's Inequality, since, within each length- $k$  block, we apply code  $\mathcal{C}_k$  and we have an average error probability of at most  $\epsilon_{b,k}$ . Therefore, by choosing  $b$  and  $k$  sufficiently large, our constructed code achieves arbitrarily close to  $\mathbf{R}$  on the non-Gaussian additive noise wireless network.

The proof of Theorem 3 holds in this new setting almost verbatim. The only difference is that we now have one rate for each source  $s \in V$  and destination set  $D \subseteq V$  and the message vector  $\mathbf{w}$  has size  $V \times \mathcal{P}(V)$ , and the expressions for the error probability in (17) must be modified. This concludes the proof of Theorem 6.

## 5 Concluding Remarks

In this work, we proved that the Gaussian noise is the worst-case noise in additive noise wireless networks. This extends the classical result that Gaussian noise is the worst-case noise for point-to-point additive noise channels, which is commonly used as a justification for the modeling of the noise in wireless systems as Gaussian noise. Thus, we provide formal evidence that this modeling is indeed justified beyond the point-to-point setting.

It is important to highlight the fact that we prove our result by actually constructing a coding scheme that performs well on a non-Gaussian network given a coding scheme designed to perform well on an AWGN network. This is different from the mutual-information-based proof for point-to-point channels, which relies on the Channel Coding Theorem, and, thus, in random coding arguments which only provide existence guarantees. Even though we do make use of random coding arguments when we construct the outer codes in section 3.3, this is done mostly as a way to construct coding schemes whose error probability can be shown to tend to zero. As mentioned in section 3.3, the outer codes must be used since the union bound would only provide an outer bound on the error probability of the form  $b\epsilon_{b,k}$ , and we do not have any guarantees on the rate of decay of  $\epsilon_{b,k}$  as  $b$  and  $k$  increase. Thus, it seems reasonable to think that given a sequence of coding schemes for an AWGN network whose error probability decays fast to 0, no outer code will be necessary.

Another important point about the techniques we introduce is that the only information about the actual noise distributions that is required for the coding scheme construction are the mean and the variance. This means that, given a wireless network with *unknown* noise distributions where only the mean and variance can be measured, it is possible to construct a sequence of coding schemes that achieves the capacity of the corresponding AWGN network.

Finally, we notice that the result in Theorem 3 is interesting in itself, since it implies that the capacity region when we restrict ourselves to coding schemes finite reading precision and allow the precision to tend to infinity along the sequence of coding schemes is equal to the unrestricted capacity. In fact, it is not difficult to change the proof of the theorem in order to prove that  $C^{(m)}$ , the capacity region when we restrict ourselves to coding schemes where only  $m$  bits after the decimal point are available, converges to the unrestricted capacity region  $C$ , as  $m \rightarrow \infty$ . Since in any practical wireless system the analog received signals must go through an ADC, this result essentially implies that by increasing the

resolution of the ADCs used in a wireless network, the capacity region of the practical system is indeed approaching the capacity region of the usual infinite-precision models used in the study of wireless networks.

## 6 Acknowledgements

The authors would like to thank Professors Gennady Samorodnitsky and Aaron Wagner from Cornell University for helpful discussions.

The research of A. S. Avestimehr and I. Shomorony was supported in part by the NSF CAREER award 0953117, NSF Grant CCF-1144000, U.S. Air Force Young Investigator Program award FA9550-11-1-0064, and NSF TRUST Center.

## A Appendix

**Lemma 4.** *Let  $\lambda$  denote the Lebesgue measure. Then, for any convex set  $S$ ,  $\lambda(\partial S) = 0$ .*

*Proof.* Consider any point  $p \in \partial S$ . Clearly,  $p \notin S^\circ$ , and by the Supporting Hyperplane Theorem [9], there exists a hyperplane that passes through  $p$  and contains  $S$  in one of its closed half-spaces. Let  $H$  be such a closed half-space. Since  $H$  is closed, it is clear that  $\partial S \subseteq H$ . Then, for any closed ball  $B_\epsilon(p)$  centered at  $p$ , it is clear that

$$\frac{\lambda(B_\epsilon(p) \cap \partial S)}{\lambda(B_\epsilon(p))} \leq \frac{\lambda(B_\epsilon(p) \cap H)}{\lambda(B_\epsilon(p))} = 1/2.$$

By Lebesgue's Density Theorem, the set

$$P = \left\{ p \in \partial S : \liminf_{\epsilon \rightarrow 0} \frac{\lambda(B_\epsilon(p) \cap \partial S)}{\lambda(B_\epsilon(p))} < 1 \right\}$$

should have Lebesgue measure zero. But since  $P = \partial S$ , we conclude that  $\lambda(\partial S) = 0$ .  $\square$

**Lemma 3.** *Suppose  $Y$  is a random variable with density  $f$ . Let  $\tilde{Y}_m = \lfloor Y \rfloor_m + U_m$ , where  $U_m$  is uniformly distributed in  $(-2^{-m-1}, 2^{-m-1})$  and independent from  $Y$ . Then each  $\tilde{Y}_m$  has a density  $f_m$ , and  $f_m$  converges pointwise almost everywhere to  $f$ .*

*Proof.* Since the density of  $U$   $(-2^{-m-1}, 2^{-m-1})$  is  $g(x) = 2^m \mathbf{1}\{x \in (-2^{-m-1}, 2^{-m-1})\}$ ,  $\tilde{Y}_m$  will have a density that can be written, for almost all  $y$ , as

$$\begin{aligned} f_m(y) &= E[g(y - \lfloor Y \rfloor_m)] = 2^m E[\mathbf{1}\{y - \lfloor Y \rfloor_m \in (-2^{-m-1}, 2^{-m-1})\}] \\ &= 2^m \Pr[y - \lfloor Y \rfloor_m \in (-2^{-m-1}, 2^{-m-1})] \\ &= 2^m \Pr[\lfloor Y \rfloor_m \in (y - 2^{-m-1}, y + 2^{-m-1})] \\ &= 2^m \Pr[\lfloor 2^m Y \rfloor \in (y2^m - 1/2, y2^m + 1/2)] \\ &= 2^m \Pr[2^m Y \in (\lceil y2^m - 1/2 \rceil, \lceil y2^m + 1/2 \rceil)] \\ &= 2^m \Pr[Y \in (2^{-m} \lceil y2^m - 1/2 \rceil, 2^{-m} \lceil y2^m + 1/2 \rceil)] \\ &= 2^m \int_{a_m}^{b_m} f(x) dx, \end{aligned} \tag{19}$$

where  $a_m = 2^{-m} \lceil y2^m - 1/2 \rceil$  and  $b_m = 2^{-m} \lceil y2^m + 1/2 \rceil$ . Notice that we can write  $b_m = a_m + 2^{-m}$ . Moreover, we have that

$$y - 2^{-(m+1)} \leq a_m < y + 2^{-(m+1)}, \tag{20}$$

from which we have  $a_m \rightarrow y$  as  $m \rightarrow \infty$ . If we let  $F(y)$  be the cdf of  $Y$ , then (19) can be written as

$$\frac{F(b_m) - F(a_m)}{2^{-m}} = \frac{F(a_m + 2^{-m}) - F(a_m)}{2^{-m}} \triangleq q_m. \quad (21)$$

Our goal is to show that  $q_m$  converges to  $f(y)$  as  $m \rightarrow \infty$  for almost all  $y$ . Since by assumption  $Y$  has an absolutely continuous distribution,  $F(y)$  is differentiable almost everywhere, so it suffices to show that  $q_m$  converges to  $f(y)$  as  $m \rightarrow \infty$  wherever  $F(y)$  is differentiable and the derivative is  $f(y)$ . Thus, we focus on a  $y$  where  $F'(y) = f(y)$ . Suppose by contradiction that  $q_m$  does not converge to  $f(y)$ . Then there must be an  $\epsilon > 0$  and a subsequence  $\{q_{m_i}\}_{i=1}^\infty$ , such that one of the following

$$q_{m_i} > f(y) + \epsilon \quad (22)$$

$$q_{m_i} < f(y) - \epsilon \quad (23)$$

holds for all  $i \geq 1$ . Suppose wlog that we have a subsequence  $\{q_{m_i}\}_{i=1}^\infty$  for which (22) holds for all  $i \geq 1$ . We will now pick a further subsequence of  $\{q_{m_i}\}_{i=1}^\infty$  in the following way. First, we choose  $K \in \mathbb{Z}_+$  large enough so that  $f(y)/K < \epsilon$ , and we define  $K$  subsets of  $\{1, 2, \dots\}$  as

$$S_j = \left\{ i \geq 1 : y - 2^{-(m_i+1)} + \frac{j-1}{K} 2^{-m_i} \leq a_{m_i} < y - 2^{-(m_i+1)} + \frac{j}{K} 2^{-m_i} \right\},$$

for  $j = 1, 2, \dots, K$ . From (20), the sets  $S_1, \dots, S_K$  partition  $\{1, 2, \dots\}$ , and we must be able to find some  $S_j$  that is infinite. Suppose  $|S_t| = \infty$ . Then we have a subsequence  $\{q_{m_i}\}_{i \in S_t}$ , which we re-index as  $\{q_{\ell_i}\}_{i=1}^\infty$ . For each of the elements in this subsequence we have

$$\begin{aligned} q_{\ell_i} &= \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(a_{\ell_i})}{2^{-\ell_i}} = \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{2^{-\ell_i}} + \frac{F(y) - F(a_{\ell_i})}{2^{-\ell_i}} \\ &= \frac{a_{\ell_i} + 2^{-\ell_i} - y}{2^{-\ell_i}} \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + \frac{y - a_{\ell_i}}{2^{-\ell_i}} \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}} \\ &\stackrel{(ii)}{\leq} \frac{2^{-\ell_i}(1 + t/K - 1/2)}{2^{-\ell_i}} \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + \frac{2^{-\ell_i}(1/2 - (t-1)/K)}{2^{-\ell_i}} \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}} \\ &= (t/K + 1/2) \frac{F(a_{\ell_i} + 2^{-\ell_i}) - F(y)}{a_{\ell_i} + 2^{-\ell_i} - y} + (1/2 - (t-1)/K) \frac{F(y) - F(a_{\ell_i})}{y - a_{\ell_i}}, \end{aligned} \quad (24)$$

where (i) follows since  $F(y)$  is non-decreasing and  $\ell_i \in S_t$ . Now, notice that the right-hand side in (24) has a limit, and, by taking the lim sup, we obtain

$$\begin{aligned} \limsup_{i \rightarrow \infty} q_{\ell_i} &\leq (t/K + 1/2)f(y) + (1/2 - (t-1)/K)f(y) \\ &= \left(1 + \frac{1}{K}\right) f(y) < f(y) + \epsilon. \end{aligned}$$

But this is a contradiction because all  $q_{m_i}$  satisfied  $q_{m_i} > f(y) + \epsilon$ , and  $\{q_{\ell_i}\}_{i=1}^\infty \subseteq \{q_{m_i}\}_{i=1}^\infty$ . We conclude that we must have

$$\lim_{m \rightarrow \infty} q_m = f(y),$$

which implies that  $f_m(y) \rightarrow f(y)$  as  $m \rightarrow \infty$ . □

## References

- [1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, 2nd edition, 2006.
- [2] A. Lapidoth. Nearest neighbor decoding for additive non-gaussian noise channels. *IEEE Transactions on Information Theory*, 42(5):1520–1529, September 1996.
- [3] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory*, 54(12):5534–5562, December 2008.
- [4] A. S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. *IEEE Transactions on Information Theory*, 57(4), April 2011.
- [5] S. H. Lim, Y. H. Kim, A. El Gamal, and S. Y. Chung. Noisy network coding. *IEEE Transactions on Info. Theory*, 57(5):3132–3152, 2011.
- [6] A. Ozgur and S. N. Diggavi. Approximately achieving gaussian relay network capacity with lattice codes. *ISIT Proceedings*, 2010.
- [7] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [8] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 3rd edition, 1995.
- [9] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.